

## Modeling of Soil Saturated Hydraulic Conductivity Using Intelligent Methods

Shokrollah Asghari<sup>1\*</sup>, Mahsa Hasanpour Kashani<sup>2</sup>

(Received: May, 2025

Accepted: September, 2025)

### Abstract

The saturated hydraulic conductivity ( $K_s$ ) of soil is essential for the transport of water, nutrients, and pollutants. Direct measurement of  $K_s$  is time-consuming, expensive, and may produce unreliable results due to soil heterogeneity and errors. Therefore, indirect methods such as regression pedotransfer functions and intelligent models are used to estimate this soil property. This study compared the performance of three intelligent models -neuro-fuzzy (NF), gene expression programming (GEP), and random forest (RF)—in estimating  $K_s$  from readily available soil variables. Data from 102 soil samples collected from agricultural lands in the Ardabil plain, including sand, silt, clay, geometric mean and geometric standard deviation of particle diameter, bulk and particle density, organic carbon,  $\text{CaCO}_3$ , and  $K_s$  were used. Eighty-two data points were used for training and 20 for testing the intelligent models. Five different combinations of readily available soil variables were selected as model inputs to estimate  $K_s$  using NF, GEP, and RF models. The results of all three intelligent methods indicated that the combination of clay and soil bulk density as input variables provided the highest accuracy and was the best model for estimating  $K_s$ . The coefficient of determination ( $R^2$ ), normalized root mean square error (NRMSE), mean error (ME), and Nash-Sutcliffe coefficient (NS) were calculated based on the test data as follows: 0.83, 0.118, 0.002  $\text{cm min}^{-1}$ , 0.81 for NF; 0.83, 0.145, 0.020  $\text{cm min}^{-1}$ , 0.71 for GEP; and 0.74, 0.151, 0.020  $\text{cm min}^{-1}$ , 0.69 for RF. The results showed that the NF model estimated  $K_s$  with high accuracy in the study area compared to the other two intelligent models, due to its low NRMSE and high NS.

**Key words:** Estimation, Gene expression, Neuro-fuzzy, Random Forest, Soil hydraulic conductivity.

Asghari Sh., Hasanpour Kashani M. 2025. Modeling of Soil Saturated Hydraulic Conductivity Using Intelligent Methods. *Applied Soil Research*. 13(4): 49-63.

1. Professor, Department of Soil Science and Engineering, Faculty of Agriculture and Natural Resources, University of Mohagheh Ardabili, Ardabil, Iran.

2. Associate professor, Department of Water Engineering, Faculty of Agriculture and Natural Resources, University of Mohagheh Ardabili, Ardabil, Iran.

\* Corresponding Author Email: 1) shokrollah.asghari@gmail.com 2) shasghari@uma.ac.ir

## مدل سازی هدایت هیدرولیکی اشباع خاک با استفاده از روش های هوشمند

شکراله اصغری<sup>۱\*</sup>، مهسا حسنیپور کاشانی<sup>۲</sup>

تاریخ دریافت: ۱۴۰۴/۰۲/۱۶

تاریخ پذیرش: ۱۴۰۴/۰۶/۱۲

## چکیده

هدایت هیدرولیکی اشباع ( $K_s$ ) خاک نقش کلیدی در انتقال آب، عناصر غذایی و آلاینده‌ها در خاک دارد. اندازه‌گیری مستقیم  $K_s$  در خاک به زمان و هزینه زیادی نیاز داشته و نتایج به‌دست آمده به‌دلیل غیریکنواختی خاک و بروز خطا ممکن است چندان قابل اعتماد نباشد. بنابراین از روش‌های غیر مستقیم مانند توابع انتقالی رگرسیونی و مدل‌های هوشمند برای تخمین این متغیر دیرپافت خاک استفاده گردیده است. هدف از این پژوهش، مقایسه عملکرد سه مدل هوشمند نرو-فازی (NF)، برنامه‌ریزی بیان ژن (GEP) و جنگل تصادفی (RF) در برآورد  $K_s$  از روی ویژگی‌های زودپافت خاک بود. برای این منظور، از اطلاعات مربوط به ۱۰۲ نمونه خاک منتخب از اراضی زراعی دشت اردبیل شامل شن، سیلت، رس، میانگین و انحراف معیار هندسی قطر ذرات، جرم مخصوص ظاهری و حقیقی، کربن آلی، آهک و  $K_s$  استفاده شد. تعداد ۸۲ داده برای آموزش و ۲۰ داده برای آزمون مدل‌های هوشمند به کار گرفته شد. پنج ترکیب مختلف از متغیرهای خاک به‌عنوان ورودی مدل برای برآورد  $K_s$  توسط سه مدل NF، GEP و RF انتخاب شدند. نتایج هر سه روش هوشمند به کار رفته در تحقیق گویای آن بود که با ترکیب دو متغیر ورودی رس و جرم مخصوص ظاهری خاک، بالاترین دقت بدست آمد. مقادیر آماره‌های ضریب تبیین ( $R^2$ )، ریشه میانگین مربعات خطای نرمال شده (NRMSE)، میانگین خطا (ME) و ضریب نش-ساتکلیف (NS) براساس داده‌های آزمون برای برابر ۰/۸۳، ۰/۱۱۸،  $\text{cm min}^{-1}$ ، ۰/۰۰۲، ۰/۸۱، ۰/۱۴۵،  $\text{cm min}^{-1}$  و ۰/۷۱/۰۲۰، ۰/۷۴، ۰/۱۵۱،  $\text{cm min}^{-1}$  و ۰/۰۲۰، ۰/۶۹ و ۰/۰۲۰ به ترتیب برای بهترین مدل NF، GEP و RF در برآورد  $K_s$  محاسبه گردید. نتایج نشان داد که مدل NF در مقایسه با دو مدل هوشمند دیگر به دلیل برخورداری از NRMSE پایین و NS بالا توانست  $K_s$  را با دقت بالایی در منطقه مورد مطالعه برآورد نماید.

واژه های کلیدی: تخمین، بیان ژن، جنگل تصادفی، نرو فازی، هدایت هیدرولیکی خاک.

اصغری ش.، حسنیپور کاشانی م. ۱۴۰۴. مدل سازی هدایت هیدرولیکی اشباع خاک با استفاده از روش های هوشمند. تحقیقات کاربردی خاک. جلد ۱۳، شماره ۴. صفحه: ۴۹-۶۳.

۱- استاد، گروه علوم و مهندسی خاک، دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران.

۲- دانشیار، گروه مهندسی آب، دانشکده دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران.

\* پست الکترونیک: 1) shokrollah.asghari@gmail.com 2) shasghari@uma.ac.ir

## مقدمه

هدایت هیدرولیکی اشباع<sup>۱</sup> ( $K_s$ ) خاک یکی از ویژگی‌های مهم فیزیکی خاک است که توانایی خاک را در انتقال آب در شرایط اشباع نشان می‌دهد و در جابجایی آب و املاح و نیز فرایندهای نفوذ آب به خاک و رواناب نقش اساسی دارد (Jury and Horton, 2004). اندازه‌گیری مستقیم هدایت هیدرولیکی اشباع ( $K_s$ ) خاک در مزرعه و یا آزمایشگاه به دلیل تغییرپذیری زمانی و مکانی بسیار بالای آن، کاری خسته‌کننده و پرهزینه می‌باشد بنابراین اخیراً محققان به برآورد  $K_s$  از روی برخی ویژگی‌های آسان و کم‌هزینه خاک با استفاده از روش‌های رگرسیونی و هوشمند مبتنی بر یادگیری ماشین روی آورده‌اند. از جمله روش‌های غیرمستقیم رایج و پرکاربرد برای تخمین  $K_s$ ، مدل‌های رگرسیون خطی چندگانه<sup>۲</sup> (MLR) و شبکه‌های عصبی مصنوعی<sup>۳</sup> (ANN) می‌باشد که توسط محققان مختلف مورد استفاده قرار گرفته است (Amirabedi et al., 2019). مدل‌های رگرسیونی تنها قادر هستند روابط خطی و غیر خطی ساده بین متغیرهای ورودی (زودیافت) و خروجی (دیریافت) خاک را مورد شناسایی قرار دهند. بنابراین در سال‌های اخیر استفاده از مدل‌های هوشمند که قادر به تشخیص روابط ریاضی بسیار پیچیده بین متغیرهای زودیافت و دیریافت خاک هستند مورد استقبال پژوهشگران علوم و مهندسی خاک قرار گرفته است؛ از جمله این روش‌های هوشمند مبتنی بر یادگیری ماشین، می‌توان به مدل‌های نروفازی<sup>۴</sup> (NF)، برنامه‌ریزی بیان ژن<sup>۵</sup> (GEP) و جنگل تصادفی<sup>۶</sup> (RF) اشاره نمود.

احمدی و همکاران (Ahmadi et al., 2018) هدایت هیدرولیکی اشباع ( $K_s$ ) را از روی متغیرهای شن، سیلت، رس، کربن آلی، جرم مخصوص ظاهری، اسیدیته و هدایت الکتریکی در ۱۶۰ نمونه خاک منتخب از استان آذربایجان شرقی با استفاده مدل برنامه‌ریزی بیان ژن (GEP) و رگرسیون ریح برآورد نمودند؛ نتایج نشان داد که دقت و صحت برآورد مدل GEP به علت داشتن ضریب تبیین ( $R^2$ ) بالا و ریشه میانگین مربعات خطا<sup>۷</sup> (RMSE) پایین (به ترتیب ۰/۹۲ و ۲/۲۷ سانتی‌متر بر ساعت)، بیشتر از

رگرسیون ریح (به ترتیب ۰/۷۰ و ۳/۴۸ سانتی‌متر بر ساعت) بود. امیرعابدی و همکاران (Amirabedi et al., 2019) از مدل MLR و ANN برای برآورد  $K_s$  از روی متغیرهای زودیافت خاک شامل شن، سیلت، رس، میانگین و انحراف معیار هندسی قطر ذرات، کربن آلی، آهک، جرم مخصوص ظاهری و حقیقی در اراضی زراعی دشت اردبیل استفاده کردند؛ نتایج نشان داد مدل ANN ( $R^2=0.84$ ) در مقایسه با مدل MLR ( $R^2=0.53$ ) از دقت بالایی برخوردار بود. سینق و همکاران (Singh et al., 2020) هدایت هیدرولیکی اشباع خاک را از روی برخی متغیرهای زودیافت خاک با استفاده از چندین مدل هوشمند برآورد نموده و گزارش کردند مدل جنگل تصادفی (RF) دارای بالاترین صحت تخمین در بین مدل‌های مورد مطالعه و رس دارای بیشترین و جرم مخصوص حقیقی دارای کمترین حساسیت در بین متغیرهای زودیافت خاک از نظر برآورد  $K_s$  بودند. نتایج پژوهش فرزاد مهر و همکاران (Farzadmehr et al., 2021) بر روی ۱۵۱ نمونه خاک منتخب از بجنورد بیانگر آن بود که مدل GEP در مقایسه با توابع انتقالی مبتنی بر رگرسیون خطی و غیر خطی توانست  $K_s$  را با  $R^2$  بالا (۰/۹۱) در مقابل ۰/۸۲ و RMSE پایین (۲/۸۴) در مقابل ۴/۷۴ سانتی‌متر بر روز) از روی متغیرهای شن، سیلت، جرم مخصوص ظاهری و حقیقی برآورد نماید. بدری و دربندی (Badri and Darbandi, 2021)  $K_s$  را در ۶۰ نمونه خاک منتخب از استان آذربایجان شرقی به روش نفوذسنج گلف در صحرا اندازه‌گیری نموده و سپس به روش نروفازی آن را مدل‌سازی کردند. نتایج نشان داد مدل نروفازی با چهار متغیر ورودی شن، سیلت، رس و جرم مخصوص ظاهری با RMSE برابر با ۲/۴۵ سانتی‌متر بر ساعت و آماره نش‌ساتکلیف<sup>۸</sup> (NS) برابر با ۰/۹۶ بهترین برآورد را داشت. رحمان و همکاران (Rehman et al., 2022) گزارش کردند مدل GEP در مقایسه با مدل ANN توانست  $K_s$  را با دقت بالا در خاک‌های شنی برآورد نماید و آنالیز حساسیت نشان داد که متغیر  $D_{10}$  (قطری که ۱۰ درصد ذرات خاک از آن کوچکتر هستند) مؤثرترین متغیر در برآورد  $K_s$  بود. پیری و همکاران (Piri et al., 2023) گزارش نمودند مدل جنگل تصادفی (RF) توانست  $K_s$  را با ضریب تبیین بالا (۰/۹۸) و خطای

۵ . Gene Expression Programming(GEP)

۶ . Random Forest (RF)

۷ . Root Mean Square Error (RMSE)

۸ . Nash Sutcliffe(NS)

۱ . Saturated hydraulic conductivity ( $K_s$ )

۲ . Multiple Linear Regression (MLR)

۳ . Artificial Neural Network (ANN)

۴ . Neuro-fuzzy (NF)

جغرافیایی ۲۳۰۰۰۰ تا ۲۹۰۰۰۰ کیلومتر و عرض جغرافیایی ۴۲۰۰۰۰ تا ۴۲۷۰۰۰۰ کیلومتر بر مبنای UTM) استفاده گردید (Amirabedi *et al.*, 2019). توزیع اندازه ذرات به روش هیدرومتری (Gee and Or, 2002) و کلاس بافت خاک در سیستم آمریکایی (شکل ۱)، میانگین ( $d_g$ ) و انحراف استاندارد ( $\sigma_g$ ) هندسی قطر ذرات خاک با استفاده از روابط ارائه شده توسط شیرازی و بورسما (Shirazi and Boersma, 1984) تعیین شدند. کربن آلی به روش والکلی-بلک (Nelson and Sommers, 1982)، جرم مخصوص ظاهری به روش استوانه دست‌نخورده (Blake and Hartge, 1986)، جرم مخصوص حقیقی به روش پیکنومتر و آهک به روش تیتراسیون با اسید و باز به دست آمد. هدایت هیدرولیکی اشباع ( $K_s$ ) به روش بار ثابت (خاک‌های درشت بافت) و بار افتان (خاک‌های ریزبافت) در استوانه‌های استیل اندازه‌گیری شد (Klute & Dirksen, 1986). آماره‌های توصیفی برخی متغیرهای خاک به کار رفته در این تحقیق و نیز ضرایب همبستگی بین آن‌ها به ترتیب در جداول ۱ و ۲ ارائه شده است؛ مشاهده می‌شود در بین متغیرهای مورد مطالعه،  $K_s$  در مقایسه با سایر متغیرهای خاک از ضریب تغییرات بسیار بالایی (۸۶/۳۱ درصد) برخوردار است (جدول ۱). این نکته اهمیت تخمین  $K_s$  از روی ویژگی‌های زودبافت خاک به جای اندازه‌گیری مستقیم آن در مزرعه یا آزمایشگاه را نشان می‌دهد. همچنین شکل ۱ نشان می‌دهد که خاک‌های مورد استفاده دارای تنوع بافتی نسبتاً خوبی هستند (۵۳ نمونه لوم، ۲۶ نمونه لوم رسی، ۱۵ نمونه لوم شنی، ۶ نمونه لوم سیلتی و ۲ نمونه رسی) که از نظر آموزش و یادگیری مدل‌ها حائز اهمیت است.

## ۲- مدل سازی داده‌ها

در این پژوهش در ادامه تحقیق امیرعابدی و همکاران (Amirabedi *et al.*, 2019) و نیز به منظور مقایسه دقیق و بهتر نتایج با خروجی مدل‌های به کار رفته در تحقیقات قبلی، از همان پنج ترکیب متفاوت متغیرهای ورودی شامل شن، سیلت، رس، جرم مخصوص ظاهری و حقیقی،  $d_g$  و  $\sigma_g$  مورد استفاده در مدل‌های رگرسیونی (جدول ۳) توسط نامبردگان، برای برآورد هدایت هیدرولیکی اشباع ( $K_s$ ) خاک با استفاده از سه مدل هوشمند نروفازی (NF)، برنامه‌ریزی بیان ژن (GEP) و جنگل تصادفی (RF) استفاده گردید.

مطلق میانگین پایین (۰/۰۰۱۹) از روی متغیرهای شن، سیلت، رس، اسیدیته، هدایت الکتریکی، کربن آلی و آهک در خاک‌های منتخب از منطقه هیرمند برآورد نماید. تان و همکاران (Tan *et al.*, 2023) از سه مدل هوشمند جنگل تصادفی، درخت تصمیم‌گیری و شبکه عصبی مصنوعی برای برآورد  $K_s$  در موانع خاکی فشرده محل‌های دفن زباله استفاده کردند؛ نتایج نشان داد که مدل جنگل تصادفی دارای بالاترین دقت پیش‌بینی نسبت به دو مدل دیگر بود. همچنین متغیرهای زودبافت شامل نسبت پوکی بعد از فشرده‌گی، مقادیر ذرات ریز، چگالی ویژه، درجه اشباع بعد از فشرده‌گی و شاخص خمیرایی به ترتیب نزولی مهمترین متغیرهای اثرگذار از نظر برآورد  $K_s$  بودند. زانگ و زانگ (Zhang and Zhang, 2024) گزارش کردند که مدل GEP توانست هدایت هیدرولیکی اشباع خاک را از روی برخی ویژگی‌های زودبافت خاک مانند نسبت پوکی، اندازه موثر ( $d_{10}$ ) و میانگین اندازه ذره ( $d_{50}$ ) با دقت بالاتری ( $R^2$  برابر با ۰/۹۴ و RMSE برابر با ۰/۰۰۴۸) نسبت به سایر مدل‌های هوشمند و روابط ریاضی تخمین بزند.

بررسی منابع نشان داد، اگرچه از برخی مدل‌های هوشمند برای برآورد هدایت هیدرولیکی اشباع ( $K_s$ ) از روی ویژگی‌های زودبافت خاک توسط برخی از محققان در داخل و خارج کشور استفاده شده است، ولی براساس اطلاعات در دسترس، از مدل‌های نروفازی (NF)، برنامه‌ریزی بیان ژن (GEP) و جنگل تصادفی (RF) به‌طور همزمان برای برآورد  $K_s$  خاک در اراضی استان اردبیل استفاده نشده است. بنابراین هدف اصلی پژوهش حاضر، ارزیابی و مقایسه دقت مدل‌های NF، GEP و RF در برآورد  $K_s$  خاک‌های زراعی دشت اردبیل با استفاده از آماره‌های ضریب تبیین ( $R^2$ )، ریشه میانگین مربعات خطای نرمال شده (NRMSE)، میانگین خطا (ME) و نش ساتکلیف (NS) بود.

## مواد و روش‌ها

### ۱- منطقه مورد مطالعه و داده‌های مورد استفاده

در تحقیق حاضر از داده‌های به دست آمده از ویژگی‌های شن، سیلت، رس، جرم مخصوص ظاهری و حقیقی، کربن آلی و آهک مربوط به ۱۰۲ نمونه دست‌خورده و دست‌نخورده (با استفاده از استوانه استیل به قطر و ارتفاع ۵ سانتی‌متر) برداشته شده از افق Ap خاک‌های زراعی دشت اردبیل با مساحت حدود ۲۲۰۰ کیلومتر مربع (طول

استفاده شد. لازم به ذکر است که محاسبات مربوط به آمار توصیفی و ضریب همبستگی پیرسون (جداول ۱ و ۲) بین متغیرها و نیز اشتقاق مدل‌های رگرسیونی خطی چندگانه (MLR) بین متغیرهای زودپافت خاک با متغیر دیریافت بود (Amirabedi *et al.*, 2019).

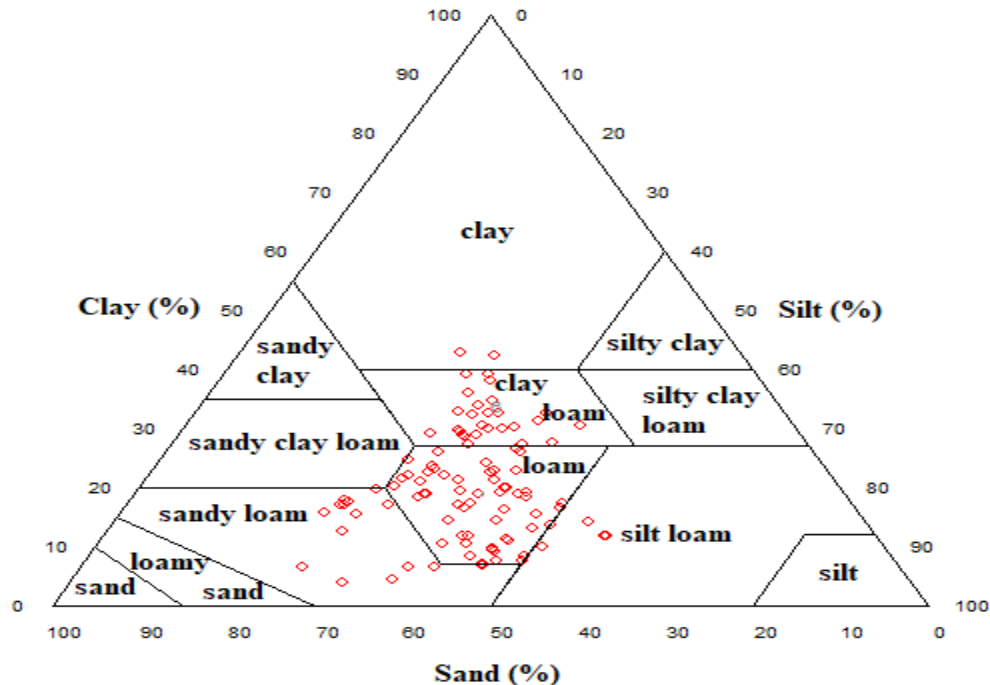
به روش گام به گام در نرم‌افزار SPSS 22 انجام شده

جدول ۱- آماره‌های توصیفی متغیرهای خاک در منطقه مورد مطالعه (n=102)

Table 1. Descriptive statistics of soil variables in the study area (n=102).

Variable	Min	Max	Mean	CV (%)
Sand (%)	24.40	68.00	42.03	21.19
Silt (%)	23.30	57.60	37.63	20.90
Clay (%)	3.80	42.90	20.33	45.39
OC (%)	0.31	1.52	0.82	31.11
BD (g cm <sup>-3</sup> )	1.04	1.70	1.36	10.77
PD (g cm <sup>-3</sup> )	2.50	2.70	2.60	1.86
d <sub>g</sub> (mm)	0.02	0.26	0.07	60.12
σ <sub>g</sub> (-)	7.48	19.41	13.61	21.90
CaCO <sub>3</sub> (%)	3.75	37.50	14.82	33.84
K <sub>s</sub> (cm min <sup>-1</sup> )	0.02	0.47	0.12	86.31

OC: Organic carbon; BD: Bulk density; PD: Particle density d<sub>g</sub>: Geometric mean diameter; σ<sub>g</sub>: Geometric standard deviation; K<sub>s</sub>: Saturated soil hydraulic conductivity; CV: Coefficient of variation.



شکل ۱- توزیع کلاس بافت در خاک‌های مورد مطالعه.

Figure 1. Distribution of textural class in the studied soils.

جدول ۲- ضریب همبستگی پیرسون (r) بین هدایت هیدرولیکی اشباع (K<sub>s</sub>) خاک و متغیرهای ورودی مدل ها (n=102).

Table 2. Pearson correlation coefficient (r) between saturated soil hydraulic conductivity (K<sub>s</sub>) and input variables of models (n=102).

Variable	Sand	Silt	Clay	dg	σ <sub>g</sub>	OC	BD	PD	CaCO <sub>3</sub>
K <sub>s</sub>	0.51**	0.33**	-0.78**	0.66**	-0.67**	-0.06	-0.20*	0.46**	-0.50**

\*\* و \* به ترتیب معنی دار در سطح احتمال ۱ و ۵ درصد. توصیف علائم متغیرها در زیر جدول ۱ آورده شده است.

Variable symbols description is available under Table 1. \*\*, \*: Significant at  $P < 0.05$  and  $P < 0.01$ , respectively.

جدول ۳- مدل های رگرسیون خطی چندگانه (MLR) هدایت هیدرولیکی اشباع (K<sub>s</sub>) خاک (Amirabedi et al., 2019).

Table 4. Multiple linear regression (MLR) models of saturated soil hydraulic conductivity (K<sub>s</sub>).

Number	Model
1	$\text{Log } K_s = 0.027 \text{ Sand} + 0.025 \text{ Silt} - 3.108$
2	$\text{Log } K_s = 0.033 \text{ Sand} + 0.028 \text{ Silt} - 0.598 \text{ BD} - 2.667$
3	$\text{Log } K_s = -0.552 \text{ BD} - 0.03 \text{ Clay} + 0.323$
4	$\text{Log } K_s = 0.013 \text{ Sand} + 1.555 \text{ PD} - 0.02 \text{ CaCO}_3 - 5.307$
5	$\text{Log } K_s = -0.534 \text{ BD} + 6.574 \text{ dg} - 0.056 \sigma_g + 0.202$

توصیف متغیرها در زیر جدول ۱ آورده شده است.

Variables description is available under Table 1.

الگوریتم برنامه ریزی بیان ژن (GEP) یک مدل توسعه یافته مشابه برنامه ریزی ژنتیک و الگوریتم ژنتیک است که در آن کروموزوم های خطی و ساده با طول ثابت مشابه در الگوریتم ژنتیک استفاده می شود و ساختارهای شاخه ای با اندازه ها و اشکال متفاوت مشابه با برنامه ریزی ژنتیک ترکیب می شوند. برای مطالعه بیشتر در خصوص نحوی اجرای مدل GEP می توان به تحقیق زانگ و زانگ (Zhang & Zhang, 2024) مراجعه نمود. در تحقیق حاضر برای برآورد هدایت هیدرولیکی اشباع (K<sub>s</sub>) خاک به روش برنامه ریزی بیان ژن (GEP) از نرم افزار Gene Xpro Tools 4.0 استفاده شد (Ferreira, 2006). در فرایند اجرای مدل GEP، مجموعه ای از توابع و ترمینال ها شامل توابع اصلی حسابی، مثلثاتی و توابع تعریف شده توسط کاربر براساس سعی و خطا تا دستیابی به حداکثر ضریب تبیین (R<sup>2</sup>) و حداقل ریشه میانگین مربعات خطا (RMSE) در برازش مدل به کار گرفته شد. عملگرهای ریاضی پرکاربرد در فرایند اجرای مدل شامل چهار عمل اصلی (+، -، × و /)، لگاریتم طبیعی و نپری، سینوس، کسینوس، آرکتانژانت، توان ۲، ۳ و ۵ بودند. پارامترهای مورد استفاده در اجرای مدل GEP به طور خلاصه در جدول ۴ و نیز درخت های توصیفی به دست آمده در پایان فرایند بهینه سازی برای

سیستم استنتاجی نروفازی (NF) ترکیبی از شبکه عصبی مصنوعی و منطق فازی است که بر پایه قوانین سوگنو اگر - آنگاه استوار می باشد. بخش فازی مدل NF یک رابطه ای را بین متغیرهای ورودی و خروجی برقرار می نماید که به آن تابع عضویت<sup>۱</sup> (mf) گفته می شود و ضرایب مربوط به توابع عضویت بخش فازی از طریق الگوریتم های آموزشی شبکه عصبی تعیین می گردد (Shiri et al., 2017). در تحقیق حاضر برای ایجاد مدل های نروفازی (NF) جهت برآورد هدایت هیدرولیکی اشباع (K<sub>s</sub>) خاک، از سیستم استنتاجی نروفازی موجود در جعبه ابزار نرم افزار MATLAB استفاده گردید. ترکیب بهینه ساختار مدل های نروفازی از طریق سعی و خطا و با استفاده از توابع عضویت و تکرارهای مختلف برای هر مدل به دست آمد (Ghorbani et al., 2019). به این ترتیب که توابع عضویت مثلثی (trimf)، گوسی (gaussmf) و پی شکل (pimf) به عنوان بهترین توابع عضویت لایه ورودی و تابع عضویت ثلثی (constant) و خطی (linear) به عنوان بهترین تابع عضویت لایه خروجی انتخاب شدند. همچنین تعداد تابع عضویت ۳ و تعداد تکرار بهینه ۱۰۰ تا برای اجرای پنج مدل NF به کار رفته در روش نروفازی تعیین گردید. روش اصلی آموزش در سیستم فازی، روش پس انتشار خطا می باشد.

<sup>1</sup> . Membership Function

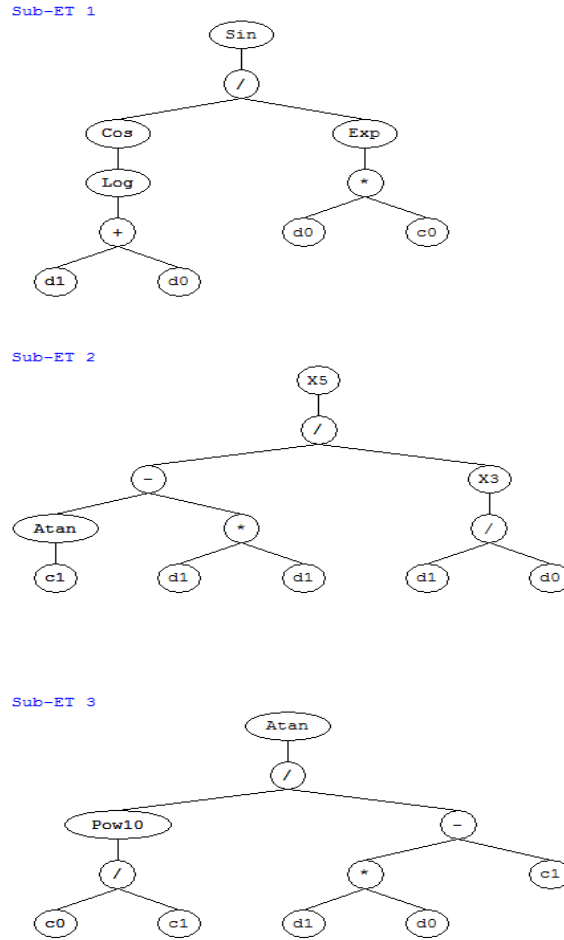
می‌شود. تعداد ورودی‌ها در هر درخت و نیز تعداد درختان بهینه، به روش سعی و خطا تعیین می‌شود. در تحقیق حاضر برای برآورد هدایت هیدرولیکی اشباع ( $K_s$ ) خاک به روش RF از نرم افزار Weka استفاده شد؛ بهترین اجرای مدل RF برای برآورد  $K_s$ ، با ۲ متغیر کمکی موجود در زیر مجموعه تصادفی هر گره ( $m_{tree}=2$ ) و تعداد ۷۰ تا ۱۰۰ درخت به‌دست آمد. کل داده‌ها (۱۰۲ داده) در اکسل به‌طور تصادفی به دو مجموعه آموزشی (۸۰ درصد داده‌ها) و آزمونی (۲۰ درصد داده‌ها) تقسیم گردید. سپس برای اجرای هر سه مدل نروفازی (NF)، برنامه‌ریزی بیان ژن (GEP) و جنگل تصادفی (RF) به‌کار گرفته شدند.

بهترین مدل GEP (مدل شماره ۳ با دو متغیر ورودی رس و جرم مخصوص ظاهری) در شکل ۲ آورده شده است. مدل جنگل تصادفی (RF) اولین بار توسط بریمن (Breiman, 2001) ارائه شد. الگوریتم RF یک الگوریتم انعطاف‌پذیر، ساده و سهل برای یادگیری ماشین می‌باشد. جنگل تصادفی یک الگوریتم یادگیری گروهی می‌باشد که مسائل رگرسیونی با ابعاد بزرگ را مدیریت می‌کند. این مدل یک روش گروهی بر پایه درختی است که در آن همه درختان به مجموعه‌ای از متغیرهای تصادفی وابسته هستند و جنگل از درختان رگرسیونی زیاد واقع در کنار یکدیگر رشد می‌یابد (Singh et al., 2020). در مدل RF برای هر درخت یک زیر مجموعه‌ای از متغیرهای ورودی اعمال

جدول ۴- مقادیر پارامترهای مدل برنامه‌ریزی بیان ژن (GEP).

Table 4. The values of gene expression programming (GEP) model parameters.

General adjustment	Values	Genetic factors	Values
Number of chromosome	30	Mutation rate	0.044
Head size	8	Inversion rate	0.1
Number of genes	3	one point recombination rate	0.3
Linking function	جمع (+)	Two point recombination rate	0.3
Fitness function error type	RMSE	Gene recombination rate	0.1



شکل ۲- درخت های توصیفی ایجاد شده توسط بهترین مدل برنامه ریزی بیان ژن (GEP)؛ d0 و d1 به ترتیب جرم مخصوص ظاهری و رس را نشان می دهد.

Figure 2. The expression trees developed by the best gene expression programming (GEP) model; d0 and d1 indicates bulk density and clay, respectively.

کارکرد ضعیف مدل می باشد (Wan et al., 2023). مقادیر مثبت ME بیش برآزشی و مقادیر منفی آن کم برآزشی مدل را نشان می دهد. اگر معیار NS برابر یک باشد بیانگر آن است که مدل سازی با دقت بالایی انجام گرفته و مقادیر اندازه گیری و برآورد شده  $K_s$  مطابقت کامل دارند. میزان NS صفر بیانگر آن است که برآوردهای مدل، دقتی برابر با دقت میانگین داده های اندازه گیری شده دارند.

(۱)

$$NRMSE = \frac{1}{Y_{\max(m)} - Y_{\min(m)}} \sqrt{\frac{\sum_{i=1}^n (Y_{i(p)} - Y_{i(m)})^2}{n}}$$

### ۳- ارزیابی مدل ها

مقایسه و ارزیابی دقت تخمین مدل های NF، GEP و RF در برآورد  $K_s$ ، با استفاده از معیارهای ضریب تبیین ( $R^2$ )، ریشه میانگین مربعات خطای نرمال شده (NRMSE)، میانگین خطا (ME) و نش ساتکلیف (NS) (Ghorbani et al., 2019) انجام گرفت. بطور کلی، هرچه مقدار  $R^2$  به یک و میزان NRMSE به صفر نزدیک باشد، دقت مدل بیشتر بوده و مقادیر اندازه گیری و برآورد شده  $K_s$  باهم برابر هستند. مقادیر NRMSE کمتر از ۰/۱ بیانگر کارکرد عالی، مقادیر ۰/۱ تا ۰/۲ بیانگر کارکرد مناسب، مقادیر ۰/۲ تا ۰/۳ بیانگر کارکرد متوسط و مقادیر بیش از ۰/۳ بیانگر

بودن آماره NS در داده‌های آزمون، مشاهده شد مدل NF شماره ۳ (جدول ۵) با دو متغیر ورودی جرم مخصوص ظاهری و رس از نظر دقت تخمین  $K_s$  از رتبه بهتری در مقایسه با سایر مدل‌های NF برخوردار بودند. مقادیر  $R^2$ ، NRMSE، ME و NS این مدل در سری آموزش به ترتیب برابر ۰/۶۵، ۰/۱۳۶، ۰/۱ cm min<sup>-1</sup> و ۰/۶۴ و در سری آزمون به ترتیب برابر ۰/۸۳، ۰/۱۱۸، ۰/۰۰۲ cm min<sup>-1</sup> و ۰/۸۱ به دست آمد (جدول ۵). مقدار صفر ME برای سری آزمون بیانگر برآورد دقیق  $K_s$  توسط مدل شماره ۳ می‌باشد و براساس آماره‌های NRMSE و NS در سری آزمون، عملکرد این مدل به ترتیب با کارکرد مناسب و بسیار خوب تلقی می‌شود؛ معادله رگرسیونی این مدل (جدول ۳) نشان می‌دهد که جرم مخصوص ظاهری و رس هر دو با علامت منفی وارد مدل گردیدند و تأثیر منفی متغیر جرم مخصوص ظاهری به علت داشتن ضریب رگرسیون بالاتر در برآورد  $K_s$  به مراتب بیشتر از متغیر رس می‌باشد. همبستگی معنی‌دار و منفی متغیرهای رس و جرم مخصوص ظاهری با  $K_s$  نیز مؤید این قضیه می‌باشد (جدول ۲). در رتبه‌های بعدی، مدل شماره ۴ با سه متغیر ورودی جرم مخصوص ظاهری، شن و  $\text{CaCO}_3$  (NRMSE = 0.159 و NS = 0.65) و مدل شماره ۱ با دو متغیر ورودی شن و سیلت (NRMSE = 0.174 و NS = 0.58) توانستند  $K_s$  را با دقت نسبتاً خوبی برآورد نمایند (جدول ۵). بنابراین در مجموع می‌توان بیان نمود که از میان پنج مدل نروفازی فوق، مدل شماره ۳ به علت داشتن دقت بالا و خطای پایین و نیز داشتن تعداد متغیر ورودی کم در مقایسه با مدل شماره ۴، برای برآورد  $K_s$  ترجیح داده می‌شود. همچنین مشاهده گردید، استفاده از روش نروفازی با دو متغیر ورودی جرم مخصوص ظاهری و رس در مقایسه با مدل‌های رگرسیونی و شبکه عصبی مصنوعی (Amirabedi et al., 2013) به علت کاهش RMSE و افزایش  $R^2$  (سری آزمون) باعث بهبود تخمین  $K_s$  شد. این نتایج با یافته‌های بدری و دربندی (Badri and Darbandi, 2021) نیز مطابقت دارد که گزارش کردند بهترین مدل نروفازی ( $R^2$  برابر ۰/۹۶ و RMSE برابر ۰/۷۵ cm min<sup>-1</sup>) با ترکیب متغیرهای ورودی جرم مخصوص ظاهری، رس، شن و سیلت در برآورد  $K_s$  به دست آمد.

$$ME = \frac{\sum_{i=1}^n Y_{i(p)} - Y_{i(m)}}{n} \quad (2)$$

$$NS = 1 - \frac{\sum_{i=1}^n (Y_{i(p)} - Y_{i(m)})^2}{\sum_{i=1}^n (Y_{i(m)} - \bar{Y}_m)^2} \quad (3)$$

در معادله‌های ۱، ۲ و ۳،  $Y_{i(p)}$ ،  $\bar{Y}_m$ ،  $Y_{i(m)}$  و  $Y_{\max(m)}$  و  $Y_{\min(m)}$  به ترتیب مقادیر اندازه‌گیری، برآورد شده توسط مدل، میانگین، حداکثر و حداقل مقادیر اندازه‌گیری شده  $K_s$  و  $n$  تعداد داده‌ها برای  $K_s$  در هر مدل می‌باشد. از میان مدل‌های ایجاد شده برای برآورد  $K_s$ ، مناسب‌ترین مدل NF، GEP و RF با ترکیب متغیرهای ورودی در داده‌های آزمون بر اساس معیارهای آماری مذکور تعیین شد. آزمون مدل‌ها در تشخیص دقت مدل به ازای داده‌هایی که در آموزش مدل استفاده نشده‌اند به کار می‌رود.

## نتایج و بحث

### ۱- نتایج مدل‌های نروفازی (NF) برای تخمین هدایت هیدرولیکی اشباع ( $K_s$ ) خاک

به منظور برآورد هدایت هیدرولیکی اشباع ( $K_s$ ) خاک، از پنج مدل نروفازی (NF) با ترکیب متغیرهای ورودی به کار رفته در توابع رگرسیونی (جدول ۲) استفاده گردید. در جدول ۵ مقادیر معیارهای ضریب تبیین ( $R^2$ )، ریشه میانگین مربعات خطای نرمال شده (NRMSE)، میانگین خطا (ME) و ضریب نش- ساتکلیف (NS) برای ارزیابی دقت مدل‌های NF در تخمین  $K_s$  آورده شده است. با عنایت به این‌که کاربرد  $R^2$  به تنهایی شاید از نظر ارزیابی اعتبار مدل‌ها، فریبنده بوده و همیشه  $R^2$  نزدیک به یک نشان‌دهنده برآورد دقیق مقادیر تخمینی و اندازه‌گیری یک متغیر دیریافت و در نتیجه پایین بودن خطا نیست (Kozak et al., 1996)، بنابراین در تحقیق حاضر، برای رتبه‌بندی هر سه مدل هوشمند مورد استفاده (NF، GEP و RF) از نظر دقت تخمین  $K_s$ ، از معیارهای NRMSE و NS در داده‌های آزمون استفاده شد. بر اساس درجه‌بندی واسیم و همکاران (Waseem et al., 2017)، اگر مقدار آماره NS بین ۱ و ۰/۷۵ باشد عملکرد مدل بسیار خوب است. اگر مقدار NS بین ۰/۷۵ و ۰/۶۵ باشد عملکرد مدل خوب است. اگر مقدار NS بین ۰/۶۵ و ۰/۵۰ باشد عملکرد مدل قابل قبول است ولی اگر مقدار آماره NS کوچکتر ۰/۵۰ باشد، عملکرد مدل نامطلوب است. بر اساس پایین بودن آماره NRMSE و بالا

جدول ۵- معیارهای ارزیابی مدل های نروفازی (NF) برای تخمین هدایت هیدرولیکی اشباع ( $K_s$ ) خاک.

Table 5. Evaluation criteria for neuro-fuzzy (NF) models for estimating saturated soil hydraulic conductivity ( $K_s$ ).

PTF number	Input variables	Training set				Testing set			
		NS	ME (cm min <sup>-1</sup> )	NRMSE	R <sup>2</sup>	NS	ME (cm min <sup>-1</sup> )	NRMSE	R <sup>2</sup>
1	Sand, Silt	0.62	0.000	0.140	0.62	0.58	0.012	0.174	0.65
2	Sand, Silt, BD	0.85	0.000	0.088	0.85	0.51	0.004	0.188	0.54
3	BD, Clay	0.64	0.000	0.136	0.65	<b>0.81</b>	<b>0.002</b>	<b>0.118</b>	<b>0.83</b>
4	BD, Sand, CaCO <sub>3</sub>	0.52	0.000	0.158	0.52	0.65	0.004	0.159	0.66
5	BD, d <sub>g</sub> , σ <sub>g</sub>	0.95	0.000	0.052	0.95	0.18	0.003	0.337	0.26

توصیف متغیرها در زیر جدول ۱ آورده شده است؛ R<sup>2</sup>: ضریب تبیین، ME: میانگین خطا، RMSE: مجذور میانگین مربعات خطا، NS: ضریب نش-ساتکلیف.

R<sup>2</sup>: coefficient of determination; RMSE: root mean square error; ME: mean error; NS: Nash-Sutcliff coefficient. Variables description is available under Table 1.

خاک و افزایش رس به علت افزایش منافذ ریز و کاهش منافذ درشت در خاک، موجب کاهش هدایت هیدرولیکی اشباع خاک می گردد (Jury and Horton, 2004). بنابراین مدل شماره ۳ با دو متغیر ورودی زود یافت شامل جرم مخصوص ظاهری و رس براساس روش برنامه ریزی بیان ژن (GEP) به عنوان بهترین مدل برای برآورد هدایت هیدرولیکی اشباع ( $K_s$ ) خاک انتخاب شد. این مدل با کمترین تعداد متغیر ورودی، دارای صحت تخمین بالایی در برآورد  $K_s$  می باشد. معادله ۴ رابطه ریاضی به دست آمده برای محاسبه  $K_s$  توسط مدل GEP شماره ۳ را نشان می دهد:

$$K_s = \sin \left( \cos \left( \frac{\log(\text{Clay} + \text{BD})}{\text{EXP}(1.179\text{BD})} \right) \right) + \left( + \frac{-1.470 - \text{Clay}^2}{\text{BD}^3} \right)^5 + \arctan \left( \frac{1.259}{\text{BDClay} - 4.014} \right) \quad (4)$$

در رتبه های بعدی تخمین  $K_s$ ، مدل GEP شماره ۱ با دو متغیر ورودی شن و سیلت (NRMSE = 0.182 و NS = 0.54) و مدل شماره ۴ با سه متغیر ورودی جرم مخصوص ظاهری، شن و CaCO<sub>3</sub> (NRMSE = 0.186 و NS = 0.47) قرار دارد (جدول ۶). همانند مدل نروفازی، استفاده از روش هوشمند GEP در مقایسه با مدل های رگرسیونی و شبکه عصبی مصنوعی (Amirabedi et al., 2013) به علت کاهش RMSE و افزایش R<sup>2</sup> (سری آزمون) باعث بهبود تخمین  $K_s$  در خاک های زراعی دشت اردبیل گردید. نتایج تحقیق

## ۲- نتایج مدل های برنامه ریزی بیان ژن (GEP) برای تخمین هدایت هیدرولیکی اشباع ( $K_s$ ) خاک

پنج مدل برنامه ریزی بیان ژن (GEP) با همان ترکیب متغیرهای ورودی مدل های رگرسیون خطی چندگانه (جدول ۳) برای برآورد هدایت هیدرولیکی اشباع ( $K_s$ ) خاک ایجاد شد. جدول ۶ مقادیر ضریب تبیین (R<sup>2</sup>)، ریشه میانگین مربعات خطای نرمال شده (NRMSE)، میانگین خطا (ME) و ضریب نش-ساتکلیف (NS) را برای مقایسه دقت مدل های GEP در برآورد  $K_s$  نشان می دهد. برای رتبه بندی عملکرد مدل های GEP در برآورد  $K_s$  نیز از همان ملاک پایین بودن NRMSE و بالا بودن NS در ارزیابی ها استفاده شد. مشاهده می شود مدل GEP شماره ۳ در داده های آزمون (جدول ۶) با دو متغیر ورودی جرم مخصوص ظاهری و رس از نظر برآورد  $K_s$  دارای دقت تخمین زیاد (NRMSE پایین و NS بالا) نسبت به سایر مدل های GEP بودند؛ مقادیر R<sup>2</sup>، NRMSE، ME و NS این مدل برای داده های آموزشی به ترتیب برابر ۰/۵۹، ۰/۱۴۶، ۰/۱۴۳ و ۰/۰۰۲- و برای داده های آزمون به ترتیب برابر ۰/۸۳، ۰/۱۴۵، ۰/۰۲ و ۰/۷۱ محاسبه گردید (جدول ۶). مقدار مثبت ME برای داده های آزمون بیانگر برآورد اندکی بیش از مقدار واقعی  $K_s$  توسط مدل شماره ۳ می باشد ولی با این حال براساس آماره های NRMSE و NS در سری آزمون، عملکرد این مدل به ترتیب با کارکرد مناسب و بسیار خوب ارزیابی می شود (Waseem et al., 2017). معادله رگرسیونی این مدل (جدول ۳) نشان می دهد که جرم مخصوص ظاهری و رس هر دو با علامت منفی وارد مدل گردیدند. در توجیه این قضیه می توان چنین گفت که افزایش جرم مخصوص ظاهری به دلیل بالا بردن فشردگی

حاضر با یافته‌های احمدی و همکاران (Ahmadi *et al.*, 2018) در یک راستا می‌باشد که گزارش کردند بهترین مدل GEP (سری آزمون) از نظر برآورد  $K_s$  (RMSE = 0.038) خاک تعیین گردید.

حاضر با یافته‌های احمدی و همکاران (Ahmadi *et al.*, 2018) در یک راستا می‌باشد که گزارش کردند بهترین مدل GEP (سری آزمون) از نظر برآورد  $K_s$  (RMSE = 0.038) خاک تعیین گردید.

جدول ۶- معیارهای ارزیابی مدل‌های برنامه‌ریزی بیان ژن (GEP) برای تخمین هدایت هیدرولیکی اشباع ( $K_s$ ) خاک.

Table 6. Evaluation criteria gene expression programming (GEP) models for estimating saturated soil hydraulic conductivity ( $K_s$ ).

PTF number	Input variables	Training set				Testing set			
		NS	ME (cm min <sup>-1</sup> )	NRMSE	R <sup>2</sup>	NS	ME (cm min <sup>-1</sup> )	NRMSE	R <sup>2</sup>
1	Sand, Silt	0.57	-0.001	0.149	0.58	0.54	0.009	0.182	0.57
2	Sand, Silt, BD	0.65	0.002	0.135	0.66	0.41	0.003	0.207	0.48
3	BD, Clay	0.59	-0.002	0.146	0.59	<b>0.71</b>	<b>0.020</b>	<b>0.145</b>	<b>0.83</b>
4	BD, Sand, CaCO <sub>3</sub>	0.29	-0.001	0.191	0.30	0.47	0.013	0.196	0.53
5	BD, d <sub>g</sub> , σ <sub>g</sub>	0.73	-0.001	0.118	0.73	0.21	0.018	0.240	0.47

توصیف متغیرها در زیر جدول ۱ آورده شده است؛ R<sup>2</sup>: ضریب تبیین، ME: میانگین خطا، RMSE: مجذور میانگین مربعات خطا، NS: ضریب نش-ساتکلیف.

R<sup>2</sup>: coefficient of determination; RMSE: root mean square error; ME: mean error; NS: Nash-Sutcliff coefficient. Variables description is available under Table 1.

مقدار مثبت ME برای سری آزمون بیانگر برآورد بیشتر از مقدار واقعی  $K_s$  توسط مدل RF شماره ۳ می‌باشد و مطابق آماره‌های NRMSE و NS در سری آزمون، عملکرد این مدل به ترتیب با کارکرد مناسب و بسیار خوب ارزیابی می‌شود (Waseem *et al.*, 2017). در رتبه‌های بعدی، مدل جنگل تصادفی شماره ۲ با سه متغیر ورودی شن، سیلت و جرم مخصوص ظاهری (NRMSE = 0.161 و NS = 0.67) و مدل RF شماره ۱ با دو متغیر ورودی شن و سیلت (NRMSE = 0.167 و NS = 0.45) توانستند  $K_s$  را با دقت نسبتاً خوبی برآورد نمایند (جدول ۷). مدل‌های RF شماره ۴ و ۵ (جدول ۷) به علت داشتن مقادیر آماره NS کوچکتر از ۰/۵، دارای عملکرد ضعیف از نظر برآورد  $K_s$  بودند. نتایج پژوهش پیری و همکاران (Piri *et al.*, 2023) نشان داد که مدل جنگل تصادفی در مقایسه با روش زمین‌آمار و روش چاید توانست هدایت هیدرولیکی اشباع خاک را با دقت بیشتری (R<sup>2</sup> = 0.98) از روی ویژگی‌های زودیافت خاک شامل بافت، آهک و کربن آلی برآورد نماید. سینق و همکاران (Singh *et al.*, 2020) نیز نتایج مشابهی را در خصوص دقت بالای مدل جنگل تصادفی در برآورد  $K_s$  از روی ویژگی‌های زودیافت خاک گزارش نموده‌اند.

### ۳- نتایج مدل‌های جنگل تصادفی (RF) برای تخمین هدایت هیدرولیکی اشباع ( $K_s$ ) خاک

پنج مدل جنگل تصادفی (RF) با ترکیب متغیرهای ورودی استفاده شده در توابع رگرسیونی (جدول ۳) برای تخمین هدایت هیدرولیکی اشباع ( $K_s$ ) خاک طراحی گردید. در جدول ۷ مقادیر ضریب تبیین (R<sup>2</sup>)، ریشه میانگین مربعات خطای نرمال شده (NRMSE)، میانگین خطا (ME) و ضریب نش-ساتکلیف (NS) برای ارزیابی عملکرد مدل‌های RF از نظر دقت پیش‌بینی  $K_s$  آورده شده است. همانند دو مدل هوشمند قبلی، رتبه‌بندی مدل‌های RF از نظر دقت تخمین  $K_s$  نیز بر اساس پایین بودن آماره NRMSE و بالا بودن آماره NS انجام شد. همانند روش‌های نروفازی (جدول ۵) و برنامه‌ریزی بیان ژن (جدول ۶)، مدل RF شماره ۳ در داده‌های آزمون (جدول ۷) با دو متغیر ورودی جرم مخصوص ظاهری و رس توانست  $K_s$  را با خطای پایین و دقت بالا در مقایسه با سایر مدل‌های RF تخمین بزند. مقادیر R<sup>2</sup>، NRMSE، ME و NS این مدل برای داده‌های آموزشی به ترتیب برابر ۰/۸۹، ۰/۰۸۰، ۰/۰۰۳ cm min<sup>-1</sup> و ۰/۸۸ و برای داده‌های آزمون به ترتیب برابر ۰/۷۴، ۰/۱۵۱، ۰/۲۰ cm min<sup>-1</sup> و ۰/۶۹ تعیین گردید (جدول ۷).

جدول ۷- معیارهای ارزیابی مدل های جنگل تصادفی (RF) برای تخمین هدایت هیدرولیکی اشباع ( $K_s$ ) خاک.

Table 6. Evaluation criteria of random forest (RF) models for estimating saturated soil hydraulic conductivity ( $K_s$ ).

PTF number	Input variables	Training set				Testing set			
		NS	ME (cm min <sup>-1</sup> )	NRMSE	R <sup>2</sup>	NS	ME (cm min <sup>-1</sup> )	NRMSE	R <sup>2</sup>
1	Sand, Silt	0.45	-0.002	0.130	0.46	0.62	0.030	0.167	0.52
2	Sand, Silt, BD	0.67	0.0005	0.072	0.71	0.64	0.020	0.161	0.79
3	BD, Clay	0.88	0.0003	0.080	0.89	0.69	0.020	0.151	0.74
4	BD, Sand, CaCO <sub>3</sub>	0.31	-0.002	0.189	0.39	0.48	0.021	0.195	0.76
5	BD, d <sub>g</sub> , σ <sub>g</sub>	0.35	-0.050	0.160	0.54	0.42	0.023	0.205	0.47

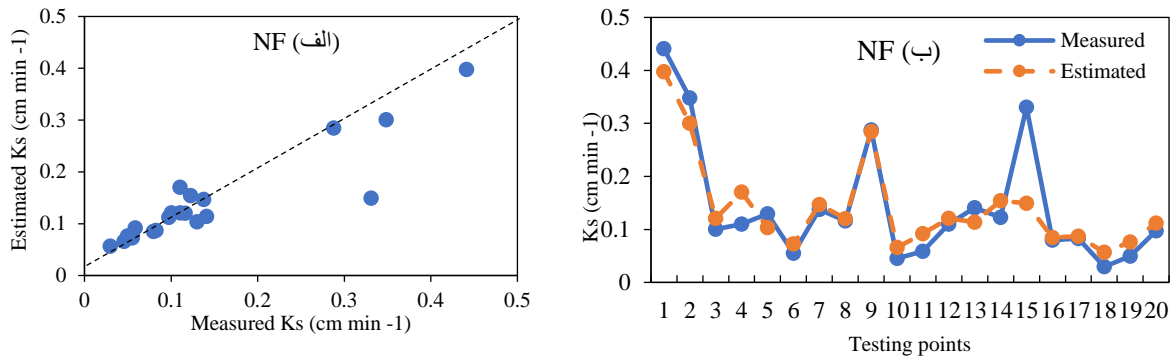
توصیف متغیرها در زیر جدول ۱ آورده شده است؛ R<sup>2</sup>: ضریب تبیین، ME: میانگین خطا، RMSE: مجذور میانگین مربعات خطا، NS: ضریب نش-ساتکلیف.

R<sup>2</sup>: coefficient of determination; RMSE: root mean square error; ME: mean error; NS: Nash-Sutcliff coefficient. Variables description is available under Table 1.

۵)، بهترین مدل GEP (مدل شماره ۳ در جدول ۶) و بهترین مدل RF (مدل شماره ۳ در جدول ۷) براساس سری آزمون نشان می دهد. با توجه به این نمودارها مشاهده می شود که برای مقادیر بزرگتر هدایت هیدرولیکی اشباع خاک، خطای هر سه روش هوشمند به کار رفته، بالا می باشد که دلیل این قضیه را شاید بتوان به تعداد اندک داده های با مقادیر  $K_s$  بالا در این تحقیق نسبت داد که موجب گردید آموزش مدل ها در این محدوده ( $K_s > 0.20$  cm min<sup>-1</sup>) به خوبی انجام نگیرد. همچنین مشاهده می شود که پراکندگی داده ها اطراف خط ۱:۱ در بهترین مدل NF نسبت به بهترین مدل های GEP و RF کمتر است. بنابراین می توان چنین استنباط نمود که در این تحقیق، مدل نروفازی در برآورد  $K_s$  از روی متغیرهای زودیافت خاک به علت استفاده از روابط غیر خطی پیچیده به مراتب موفق تر از دو مدل هوشمند دیگر عمل نموده است. در پژوهش سینق و همکاران (Singh et al., 2020) و تان و همکاران (Tan et al., 2023) مدل های مبتنی بر جنگل تصادفی در مقایسه با شبکه عصبی مصنوعی و نروفازی از دقت بالایی در برآورد  $K_s$  برخوردار بودند. زانگ و زانگ (Zhang and Zhang, 2024) نیز گزارش نمودند مدل برنامه ریزی بیان ژن و مدل جنگل تصادفی قادر بودند با دقت تقریباً برابری هدایت هیدرولیکی اشباع ( $K_s$ ) خاک را از روی متغیرهای زودیافت خاک تخمین بزنند. بنابراین با عنایت به لندک بودن تعداد مطالعات انجام گرفته و نیز ضد و نقیض بودن نتایج مطالعات قبلی، برای ارزیابی بهتر و نیز مقایسه دقیق تر عملکرد مدل های هوشمند از نظر برآورد  $K_s$ ، لازم است تحقیقات بیشتری در آینده انجام بگیرد.

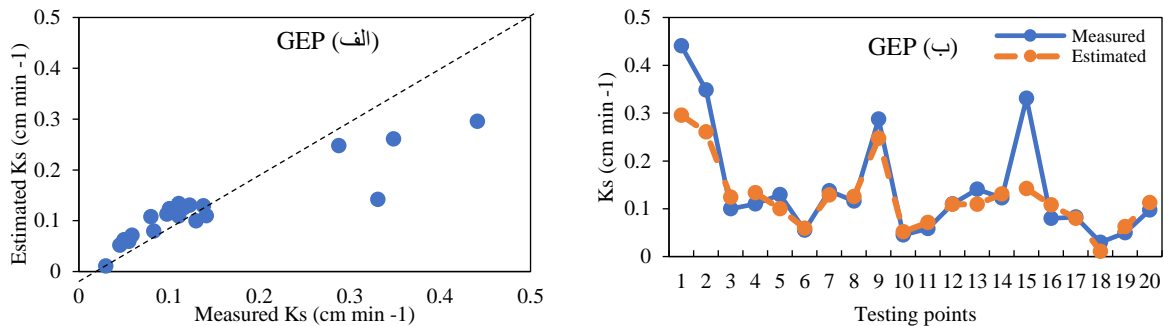
#### ۴- مقایسه عملکرد مدل های NF، GEP و RF در تخمین $K_s$

مقادیر لندازه گیری و تخمینی هدایت هیدرولیکی اشباع ( $K_s$ ) خاک بر اساس ۲۰ نقطه آزمون برای بهترین مدل نروفازی (NF)، برنامه ریزی بیان ژن (GEP) و جنگل تصادفی (RF) به ترتیب در شکل ۳، شکل ۴ و شکل ۵ (ب) نشان داده شده است. نتایج آماره های ارزیابی ریشه میانگین مربعات خطای نرمال شده (NRMSE) و ضریب نش-ساتکلیف (NS) برای بهترین مدل NF (مدل شماره ۳ در جدول ۵)، بهترین مدل GEP (مدل شماره ۳ در جدول ۶) و بهترین مدل RF (مدل شماره ۳ در جدول ۷) از نظر برآورد  $K_s$  در سری آزمون، بیانگر آن است که بهترین مدل NF به علت داشتن مقدار NS بالا (۰/۸۱) و NRMSE پایین (۰/۱۱۸) در مقایسه با بهترین مدل GEP (NRMSE = 0.145 و NS = 0.71) و بهترین مدل RF (NRMSE = 0.151 و NS = 0.69) دارای دقت بیشتر و خطای کمتر در برآورد  $K_s$  بودند. همچنین مشاهده می شود، بهترین مدل GEP در مقایسه با بهترین مدل RF (به دلیل برخورداری از NS بالاتر و NRMSE پایین تر) قادر است  $K_s$  را با دقت بهتری برآورد نماید. همچنین براساس درجه بندی واسیم و همکاران (Waseem et al., 2017) می توان گفت که در تحقیق حاضر در بین مدل های پنجگانه NF (جدول ۵)، GEP (جدول ۶) و RF (جدول ۷)، مدل شماره ۵ در هر سه روش تخمین به دلیل داشتن کوچکترین مقدار آماره NS، دارای ضعیف ترین عملکرد در برآورد هدایت هیدرولیکی اشباع ( $K_s$ ) خاک در اراضی دشت اردبیل می باشد. شکل ۳، شکل ۴ و شکل ۵ (الف) نیز به ترتیب نمودار ۱:۱ مقادیر لندازه گیری و تخمینی  $K_s$  را با استفاده از بهترین مدل NF (مدل شماره ۳ در جدول



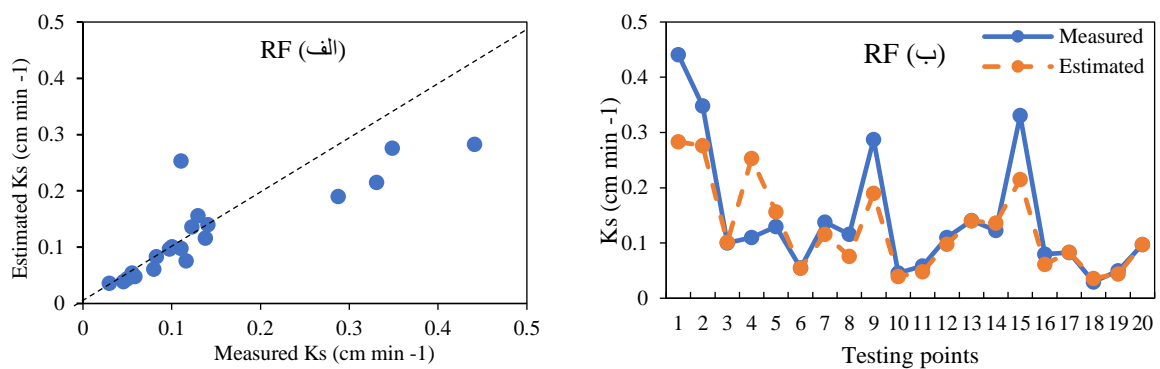
شکل ۳- مقایسه مقادیر اندازه‌گیری و برآورد شده هدایت هیدرولیکی اشباع ( $K_s$ ) خاک با استفاده از بهترین مدل نروفازی (NF) بر اساس داده‌های آزمونی ( $n=20$ )؛ اطلاعات مربوط به مدل و متغیرهای ورودی در جدول ۵ آورده شده است.

Figure 3. Comparison of estimated and measured values of soil saturated hydraulic conductivity ( $K_s$ ) using the best neuro-fuzzy (NF) model based on testing data ( $n=20$ ); Table 5 provides detailed information about model and input variables.



شکل ۴- مقایسه مقادیر اندازه‌گیری و برآورد شده هدایت هیدرولیکی اشباع ( $K_s$ ) خاک با استفاده از بهترین مدل برنامه‌ریزی بیان ژن (GEP) بر اساس داده‌های آزمونی ( $n=20$ )؛ اطلاعات مربوط به مدل و متغیرهای ورودی در جدول ۵ آورده شده است.

Figure 4. Comparison of estimated and measured values of soil saturated hydraulic conductivity ( $K_s$ ) using the best gene expression programming (GEP) model based on testing data ( $n=20$ ); Table 5 provides detailed information about model and input variables.



شکل ۵- مقایسه مقادیر اندازه‌گیری و برآورد شده هدایت هیدرولیکی اشباع ( $K_s$ ) خاک با استفاده از بهترین مدل جنگل تصادفی (RF) بر اساس داده‌های آزمونی ( $n=20$ )؛ اطلاعات مربوط به مدل و متغیرهای ورودی در جدول ۵ آورده شده است.

Figure 5. Comparison of estimated and measured values of soil saturated hydraulic conductivity ( $K_s$ ) using the best random forest (RF) model based on testing data ( $n=20$ ); Table 5 provides detailed information about model and input variables.

## نتیجه گیری کلی

از سه مدل هوشمند نروفازی (NF)، برنامه ریزی بیان ژن (GEP) و جنگل تصادفی (RF) برای تخمین هدایت هیدرولیکی اشباع ( $K_s$ ) خاک در قالب پنج ترکیب مختلف از متغیرهای زودیافت خاک شامل شن، سیلت، رس، میانگین و انحراف معیار هندسی قطر ذرات، جرم مخصوص ظاهری و آهک در خاک های زراعی دشت اردبیل استفاده شد. هدایت هیدرولیکی اشباع ( $K_s$ ) دارای همبستگی مثبت و معنی دار با شن و سیلت و همبستگی منفی و معنی دار با رس و جرم مخصوص ظاهری بود. مقایسه عملکرد مدل ها نشان داد که در هر سه روش هوشمند به کار رفته در تحقیق، مدل با دو متغیر ورودی رس و جرم مخصوص ظاهری دارای

بالاترین دقت و کارایی از نظر برآورد  $K_s$  بودند در حالی که مدل با سه متغیر ورودی جرم مخصوص ظاهری، میانگین و انحراف معیار هندسی قطر ذرات خاک دقت کمی در تخمین  $K_s$  داشت. در بین سه مدل هوشمند مورد استفاده، مدل NF به علت داشتن آماره نش ساتکلیف (NS) بالاتر و ریشه میانگین مربعات خطای نرمال شده (NRMSE) پایین تر، دارای دقت بیشتری در تخمین  $K_s$  بود. همچنین مدل GEP در مقایسه با مدل RF توانست عملکرد مناسب تری در برآورد  $K_s$  از خود نشان دهد. توصیه می گردد در تحقیقات آتی از سایر مدل های مبتنی بر یادگیری ماشین مانند ماشین بردار پشتیبان نیز در برآورد  $K_s$  در منطقه مورد مطالعه استفاده شود.

## Reference

- Amirabedi H., Asghari Sh., Mesri T., Balandeh N., and Johari E. 2019. Estimating the soil saturated hydraulic conductivity in Ardabil plain soils using artificial neural networks and regression models. *Applied Soil Research*, 7(4): 88-109. (In Persian)
- Ahmadi A., Palizvan zand P. and Palivan zand, H. 2018. Estimation of saturated hydraulic conductivity by using gene expression programming and ridge regression (A case study in East Azerbaijan province). *Iranian Journal of Soil and Water Research*, 48(5): 1087-1095. (In Persian)
- Badri Z. and Darbandi, S. 2022. Simulation of soil hydraulic conductivity using adaptive fuzzy neural inference system model (Case study of East Azarbaijan Province soils). *Water and Soil Science*, 32(2): 179-189. (In Persian)
- Blake G.R., and Hartge K.H. 1986a. Bulk density, In: Klute, A. (Ed.). *Methods of Soil Analysis. Part 1. 2nd ed. Agronomy. Monograph. 9. Madison, WI: Soil Science Society of America*; pp. 363-375.
- Breiman L. 2001. Random forests. *Machine Learning*, 45: 5-32.
- Ferreira C. 2006. *Gene Expression Programming: Mathematical modeling by an artificial intelligence*. Springer, Berlin, Heidelberg, New York, p. 478.
- Farzadmehr M., Dastourani M., Khashei Siuki A., and Jalali Moakhar V.R. 2021. Estimating the saturated hydraulic conductivity of soil using Gene expression programming method and comparing it with the pedotransfer functions. *Journal of watershed Management Research*, 11(22):155-164. (In Persian)
- Gee G.W., and Or D. 2002. Particle-size analysis. In: Dane J. H., and Topp G. C. (Eds.), *Methods of soil analysis—Part 4. Physical Methods—SSSA Book Series No. 5. Soil Science Society of America, Madison, WI*, pp. 255-293.
- Ghorbani M.A., Deo R.C., Kashani M.H., Shahabi M., and Ghorbani S. 2019. Artificial intelligence-based fast and efficient hybrid approach for spatial modeling of soil electrical conductivity. *Soil and Tillage Research*, 186: 152–164.
- Jury W., and Horton R. 2004. *Soil Physics*. John Wiley and Sons, Inc.
- Klute A., and Dirksen C. 1986. Hydraulic conductivity and diffusivity: Laboratory methods. p. 687-734. In: Klute A(ed). *Methods of Soil Analysis. Part 1, Physical and Mineralogical Methods, 2nd ed. ASA and SSSA, Madison, WI*.
- Kozak E., Pachepsky Y.A., Sokolowski S., Sokolowska Z., and Stepniewski W. 1996. A modified number-based method for estimating fragmentation fractal dimensions of soils. *Soil Science Society of America Journal*, 60: 1291-1297.

- Nelson D.W., and Sommers L.E. 1982. Total carbon, organic carbon, and organic matter. p. 539–579. In A.L. Page et al. (ed.) *Methods of Soil Analysis. Part 2*. 2<sup>nd</sup> ed. Agron. Monogr. 9. ASA and SSSA, Madison, WI.
- Piri H., Mobaraki M. and Mir M. 2023. Comparison and application of random forest, chaid and geostatistics models in predicting soil saturated hydraulic conductivity. *Journal of Ecohydrology*, 10(2): 173-185.
- Rehman Z., Khalid U., Ijaz N., Mujtaba H., Haider A., Farooq K., and Ijaz Z. 2022. Machine learning-based intelligent modeling of hydraulic conductivity of sandy soils considering a wide range of grain sizes. *Engineering Geology*, 311, 106899.
- Shirazi M.A., and Boersma L. 1984. A unifying quantitative analysis of soil texture. *Soil Science Society American Journal*, 48 (1): 142-147.
- Shiri J., Keshavarzi A., Kisi O., and Karimi S. 2017. Using soil easily measured parameters for estimating soil water capacity: Soft computing approaches. *Computers and Electronics in Agriculture*, 141: 327-339.
- Singh V.K., Kumar D., Kashyap P.S., Singh P.K., Kumar A., and Singh S.K. 2020. Modelling of soil permeability using different data driven algorithms based on physical properties of soil. *Journal of Hydrology*, 580(124223).
- Tan Y., Zhang P., Chen J., Shamet R., Nam B.H., and Pu H. 2023. Predicting the hydraulic conductivity of compacted soil barriers in landfills using machine learning techniques. *Waste Management*, 157: 357-366.
- Wan H., Qi H., and Shang S. 2023. Estimating soil water and salt contents from field measurements with time domain reflectometry using machine learning algorithms. *Agricultural Water Management*, 285, 108364.
- Waseem M., Mani N., Andiego G., and Usman, M. 2017. A review of criteria of fit for hydrological models. *International Research Journal of Engineering and Technology*, 4(11): 1765-1772.
- Zhang R., and Zhang S. 2024. Coefficient of permeability prediction of soils using gene expression programming. *Engineering Applications of Artificial Intelligence*, 128(107504).